DOI:10.11931/guihaia.gxzw202412006

三脉水丝梨叶绿体基因组特征及系统发育分析

熊霜，周富琴，王仕东，李锐，王树宝，黄媛*

（云南师范大学 生命科学学院，昆明 650500）

摘　要：三脉水丝梨（*Sycopsis triplinervia*）是金缕梅科（Hamamelidaceae）水丝梨属（*Sycopsis*）的一种常绿灌木。由于水丝梨属的系统发育地位仍存在争议且与假蚊母属（*Distyliopsis*）、蚊母树属（*Distylium*）等近缘属之间的亲缘关系尚不明确，该研究对三脉水丝梨的叶绿体基因组进行了测序与组装，并结合公共数据库中金缕梅科其他物种的叶绿体基因组数据，开展了比较基因组分析和叶绿体基因组系统发育分析。结果表明：（1）三脉水丝梨基因组大小为 159 375 bp，共编码 133 个基因，包括 8 个 rRNA 基因、37 个 tRNA 基因、87 个蛋白编码基因和 1 个假基因。（2）在其基因组中共检测到 33 个散在重复序列、39 个串联重复序列和 82 个简单重复序列（SSRs）。（3）密码子偏好以 A/U 结尾，共包含 9 个最优密码子且密码子偏好主要受自然选择的影响。（4）与近缘种相比，三脉水丝梨的叶绿体基因组整体较为保守。从水丝梨属中筛选到的 15 个高变异区域具有潜在用于分子鉴定的价值。（5）系统发育分析显示，金缕梅科为单系类群，金缕梅属（*Hamamelis*）、白缕梅属（*Parrotiopsis*）、水丝梨属、假蚊母属、波斯铁木属（*Parrotia*）和蚊母树属 6 个属形成了一个强力支持的单系分支，几个属之间具有较近的亲缘关系，其中三脉水丝梨与该分支中的其他类群互为姐妹群。然而，水丝梨属、波斯铁木属、假蚊母属和蚊母树属均非单系。该研究为金缕梅科的系统发育研究提供了重要的基础数据和参考。

关键词：三脉水丝梨，叶绿体基因组，水丝梨属，金缕梅科，基因组系统学

中图分类号：Q943　　　　　**文献标识码：A**

Analysis of chloroplast genomic characteristics and phylogeny in *Sycopsis triplinervia*

XIONG Shuang, ZHOU Fuqin, WANG Shidong, LI Rui, WANG Shubao, HUANG Yuan*

(*School of Life Sciences*, *Yunan Normal University*, Kunming 650500, China)

Abstract: *Sycopsis triplinervia* is an evergreen shrub in the genus *Sycopsis* (Hamamelidaceae). The phylogenetic position of *Sycopsis* has been a subject of debate, and its evolutionary relationships with closely related genera such as *Distyliopsis* and *Distylium* remain unclear. Here we sequenced and assembled its chloroplast genome and conducted comparative genomic and phylogenomic analyses together with publicly available Hamamelidaceae chloroplast genomes. The results were as follows: (1) The chloroplast genome of *S. triplinervia* was 159 375 bp in length and encoded 133 genes, including 8 rRNA genes, 37 tRNA genes, 87 protein-coding genes and 1 pseudogene. (2) A total of 33 interpersed repeats, 39 tandem repeats and 82 simple sequence repeats (SSRs) were identified in the genome. (3) Codon usage was biased toward A/U endings. Nine optimal codons were detected, and natural selection was determined to be the main driver of codon usage bias. (4) Overall, the chloroplast genome of *S. triplinervia* was highly conserved relative to its close relatives. Fifteen highly variable regions were identified as potential markers

---

基金项目：国家自然科学基金（32360065）。

第一作者：熊霜（1999—），硕士研究生，研究方向为植物进化生物学，（E-mail）xs1882022@163.com。

*__通信作者：__黄媛，博士，教授，主要从事植物系统与进化研究，（E-mail）huangyuanynnu@aliyun.com。

for molecular identification in *Sycopsis*. (5) Phylogenetic analyses indicated that Hamamelidaceae was monophyletic, and the six genera *Hamamelis*, *Parrotiopsis*, *Sycopsis*, *Distyliopsis*, *Parrotia*, and *Distylium* formed a strongly supported monophyletic clade. Within this clade, *S. triplinervia* was sister to the remaining taxa. However, *Sycopsis*, *Parrotia*, *Distyliopsis*, and *Distylium* were each non-monophyletic. This study provides fundamental data and a valuable reference for further phylogenetic research on Hamamelidaceae.
Keywords: *Sycopsis triplinervia*, chloroplast genome, *Sycopsis*, Hamamelidaceae, phylogenomic phylogenetics

金缕梅科（Hamamelidaceae）以常绿或落叶乔木和灌木为主，有 27 属，80~120 种，主要分布于亚洲东部，约有三分之二的种类集中在中国南部（Zhang et al., 2003; APG IV, 2016）。金缕梅科是被子植物中较为原始的类群，从白垩纪到第三纪的地层里都有发现过金缕梅科植物的化石（Zhang et al., 2003）。金缕梅科植物的原始性和复杂性，以及该科植物在外部特征、花粉形态等方面的强烈分化，使它成为植物学家探讨被子植物起源与早期分化的重要类群（张志耘，1999）。

水丝梨属（*Sycopsis*）是金缕梅科的一个小属，主要分布于中国的西南各省和印度等地（Zhang et al., 2003）。自 1818 年 R. Brown 建立了金缕梅科后，研究者从形态相似性、性状演化以及分子系统学等方面对金缕梅科进行过大量研究，建立了至少 14 个金缕梅科分类系统（Harms, 1930; Bogle & Philbrick, 1980; Endress, 1989; Li, 1997; 张志耘，1999; Zhang et al., 2003; APG IV, 2016）。不同的分类系统对水丝梨属、假蚊母属（*Distyliopsis*）、波斯铁木属（*Parrotia*）、蚊母树属（*Distylium*）以及银缕梅属（*Shaniodendron*）的分类处理不完全一致。Endress (1970) 依据形态学特征将原属于水丝梨属的 4 种植物从该属中分出另立一新属——假蚊母属，认为它与水丝梨属较疏远而与蚊母树属关系密切。《中国植物志》对中国金缕梅科的分类系统进行了修订，认为水丝梨属共包含 9 个物种，分为水丝梨亚属和后生水丝梨亚属，不支持从水丝梨属分出假蚊母属（张宏达，1979）。结合形态学和分子系统学的证据，李建华（1997）和 *Flora of China* 都认为水丝梨属仅包含 2~3 个物种，其中水丝梨和三脉水丝梨均为中国特有，从水丝梨属分出假蚊母属具有一定的合理性（Zhang et al., 2003）。APG IV 系统（2016）也支持假蚊母属的独立性，将原属于后生水丝梨亚属的 5 个物种都归到了假蚊母属。水丝梨属、假蚊母属、波斯铁木属以及蚊母树属之间能相互杂交，杂交品种的特征往往介于两亲本之间（Endress, 1989; Johnson, 2024）。因此，仅依据形态特征对水丝梨属的物种进行分类常常不够准确，对该属物种的界定应综合形态特征、群体遗传分化、系统发育关系等多重标准（刘建全，2016）。在基于叶绿体基因 *mat*K 构建的系统发育树中发现水丝梨属与蚊母树属构成姐妹类群（Li et al., 1999a）。同时，在基于核 ITS 序列的研究中发现蚊母树属与假蚊母属形成姐妹分支，而水丝梨属与波斯铁木属以及银缕梅属之间亲缘关系较近（Li et al., 1999b）。Wang 等（2022）基于叶绿体全基因组构建了金缕梅科的系统发育树，发现水丝梨（*Sycopsis sinensis*）镶嵌在蚊母树属内部。在前人的这些研究中，水丝梨属均只有水丝梨一个物种的取样，有的还缺失了假蚊母属，并没有阐明水丝梨属是否是单系，以及其与近缘属之间的系统关系。因此，需要对水丝梨属进行更广泛的取样和更多的分子系统学的证据来解决水丝梨属与假蚊母属、蚊母树属等近缘属的系统发育关系。

基于 *rbc*L、*mat*K 和 ITS 等片段序列的分子系统学研究虽能揭示一些不同分类阶元的系统发育关系，但由于片段序列所含信息位点有限且进化速率不同，其构建的系统发育树往往存在差异（陈丽琼等，2022）。为了构建更为可靠的系统发育树，需要整合更多的基因或基因组数据。随着二代测序技术的发展，叶绿体全基因组被广泛用于研究植物的系统发育。叶绿体基因组的大小一般为 107~218 kb，测序成本低，较核基因组具有保守性、遗传稳定性和

不存在基因重组等优点，更适合应用于植物系统发育和进化研究（Corriveau & Coleman, 1988; Grevich & Daniell, 2005; Ravi et al., 2008; Daniell et al., 2016）。水丝梨属中水丝梨的叶绿体基因组被详细报道过（Peng et al., 2020），但水丝梨属其余物种叶绿体基因组特征和系统发育相关分析等方面的研究成果则尚未见详细报道。

因此，本研究对三脉水丝梨（*Sycopsis triplinervia*）叶绿体基因组进行测序、组装和注释，结合公共数据库已有的金缕梅科叶绿体基因组数据，进行了比较基因组和系统发育分析，拟探讨以下科学问题：（1）三脉水丝梨的叶绿体基因组的分子结构；（2）三脉水丝梨叶绿体基因组的重复序列、SSRs 位点以及密码子偏好性；（3）基于叶绿体基因组数据构建金缕梅科系统发育树，分析三脉水丝梨的系统位置，为探讨水丝梨属与假蚊母属、波斯铁木属以及蚊母树属之间的关系提供新的分子证据。

1 材料与方法

1.1 研究材料及 DNA 提取、测序

三脉水丝梨植物分子材料采集于昭通市大关县三江口林场（103°56′20″ E、28°12′47″ N），选取生长状态良好的新鲜叶片，置于变色硅胶中进行干燥处理，样品编号为 08CS358，标本保存在中国科学院昆明植物研究所标本馆（编号：KUN1325573）。采用改良的 CTAB 法（Porebski et al., 1997）提取干燥叶片的 DNA，并对 DNA 纯度进行分析（OD260/OD280 为 1.91；浓度为 160.96 ng·μL^{-1}）。测量合格后通过超声波将提取的 DNA 破碎为 150 bp 片段构建文库来进行高通量测序，测序共获得 2.1 G raw data。测序产生的原始数据已上传至 NCBI 数据库（Submission ID: SUB14370161; BioProject ID: PRJNA1098481）。

1.2 叶绿体基因组的组装与注释

通过 fastp v0.23.2 对原始数据进行质控（Chen, 2023），随后使用 GetOrganelle v1.7.3.5.（Jin et al., 2020）软件对叶绿体基因组进行组装。组装好的序列先导入 Geneious Prime 2023 中使用 Repeat Finder 确定反向重复序列，然后再通过 NCBI 中的 blast 查找相似度最高的序列 [蚊母树（*Distylium racemosum*）] 作为参考，并使用在线软件 GeSeq（https://chlorobox.mpimp-golm.mpg.de/geseq.html）对序列进行注释（Tillich et al., 2017）。将注释好的序列在 Geneious Prime 2023 进行手动矫正，并将注释好的序列上传至 NCBI 数据库（登录号为 PP625971）。最后通过在线软件 CPGView (http://www.1kmpg. cn//cpgview)绘制三脉水丝梨叶绿体基因组的环形物理图谱。

1.3 叶绿体基因组重复序列与 SSR 分析

使用 REPuter（Kurtz et al., 2001）检测三脉水丝梨叶绿体基因组的散在重复序列（Interpersed repeat），具体参数设定如下：最小重复长度为 30 bp，汉明距离为 3。使用 MISA（Beier et al., 2017）分析三脉水丝梨叶绿体基因组的简单重复序列（Simple Sequence Repeats, SSRs），参数设置单核苷酸到六核苷酸的最小重复次数分别为 10、5、4、3、3 和 3。此外，串联重复序列（Tandam repeat）通过 Tandem Repeats Finder（Benson, 1999）进行分析。

1.4 密码子偏好性分析

使用 Geneious Prime 2023 提取三脉水丝梨叶绿体基因组的 CDS 序列，筛选出序列长度 ≥300 bp 的 CDS 序列，重复序列只保留一条，确保序列中碱基类型仅包含 A、T、C、G 且序列中间没有终止密码子。然后运行 CodonW v1.4.2 软件（Peden, 2005）计算相对同义密码子使用度（relative synonymous of condon usage, RSCU）和有效密码子数（effective number of

codon, ENC）。用 EMBOSS（https://www.bioinformatics.nl/emboss-explorer/）中的 CUSP 程序计算每条 CDS 序列的 GC 总含量以及密码子 3 个位置的 GC 含量（分别记作 GC_{all}、GC_1、GC_2 和 GC_3）。GC_1 和 GC_2 的平均值以及同义密码子第 3 位的 GC 含量分别记作 GC_{12} 和 GC_{3S}。

对计算出的 ENC 值进行排序，两端各选出 5 个基因建立高、低表达基因库，计算ΔRSCU 值。选取ΔRSCU>0.08 的密码子作为高表达密码子。结合 RSCU>1 的高频率密码子与ΔRSCU>0.08 高表达密码子得到最优密码子。

以每条序列的 ENC 和 GC_{3S} 分别为纵坐标和横坐标绘制 ENC-plot 散点图。以每条序列的 GC_3 和 GC_{12} 分别为横坐标、纵坐标绘制散点图和拟合直线，并计算相关性。利用软件 MEGA v7.0.26 计算密码子第三位每个碱基的含量（A3、T3、C3、G3），以每条 CDS 序列的 A3/(A3+T3)和 G3/(G3+C3)为坐标绘制散点图进行 PR2-plot 分析。数据可视化均由 R 包 ggplot2 v3.4.2 完成。

1.5 叶绿体基因组的比较分析

使用在线软件 CPJS draw（Li et al., 2023）比较以下 7 条金缕梅科植物的叶绿体基因组序列的 IR 边界信息，即三脉水丝梨（PP625971）、水丝梨（NC071198）、假蚊母（*Distyliopsis dunnii*, NC071205）、樟叶假蚊母（*Distyliopsis laurifolia*, NC071202）、小叶蚊母树（*Distylium buxifolium*, NC059888）、中华蚊母树（*Distylium chinense*, NC059885）和波斯铁木（*Parrotia persica*, NC071840）。使用 Geneious Prime 2023 软件中的 Mauve 插件（Darling et al., 2004）分析上述 7 条序列的共线性。以蚊母树作为参考，通过在线软件 mVISTA（Frazer et al., 2004），采用 LAGAN 运行模式分析 7 条叶绿体基因组的序列一致性。

1.6 核苷酸多态性分析

使用 DnaSP v6.0（Rozas et al., 2017）对水丝梨属三脉水丝梨和水丝梨共 4 条叶绿体基因组序列（*Sycopsis triplinervia* PP625971、*Sycopsis sinensis* NC071198、*Sycopsis sinensis* MT323104 和 *Sycopsis sinensis* MN496080）的蛋白质编码区和非编码区（包括内含子和基因间隔区）分别进行核苷酸多态性分析。

编码区的提取在 Geneious Prime 2023 软件中完成。非编码区的提取由 perl 脚本 3_extract_bed_coding_and_noncoding 完成，并手动检查提取结果。将提取的序列进行 MAFFT 比对，并将比对好的序列导入 DnaSP v6.0 进行核苷酸多态性分析。

1.7 系统发育分析

为阐明三脉水丝梨与金缕梅科其他物种的系统发育关系，从 NCBI 数据库中下载了 53 种金缕梅科植物的 60 条叶绿体全基因组序列和 3 种蕈树科（Altingiaceae）植物的 4 条叶绿体全基因组序列，加上本次组装所得的三脉水丝梨，共计 65 条序列用于系统发育分析。以蕈树科植物的叶绿体基因组为外类群，构建金缕梅科叶绿体基因组共有 CDS 系统发育树和叶绿体全基因组系统发育树。

共有 CDS 建树，以蚊母树和水丝梨为参考基因组，使用 GeSeq（https://chlorobox.mpimp-golm.mpg.de/geseq.html）在线软件（Tillich et al., 2017）对 65 条叶绿体基因组序列进行批量注释。使用 python 脚本提取出 65 条序列共有的 78 个 CDS，然后利用 MAFFT v7.525 软件（Katoh et al., 2002）对每个 CDS 进行比对，最后以蕈树科植物的叶绿体基因组为外类群，使用 IQ tree v2.1.4 软件（Nguyen et al., 2015）构建金缕梅科叶绿体基因组共有 CDS 系统发育树。

叶绿体全基因组建树，将 65 条序列使用 MAFFT v7.525 软件进行比对构建成一个超级矩阵，然后使用 Trimal v1.4 软件（Capella-Gutiérrez et al., 2009）去除矩阵中较差的区域，参数设置为“gt 为 0.75，resoverlap 为 0.90，seqoverlap 为 0.90”。使用 IQ tree v2.1.4 软件，

参数设置同上，基于 TVM+F+I+R3 最佳替代模型，构建系统发育树。

在 IQ tree v2.1.4 中构建的金缕梅科叶绿体基因组共有 CDS 系统发育树和叶绿体全基因组系统发育树，均通过重复测试 10 000 次的 SH 近似似然比（SH-aLRT）和超快自展法（UFboot）评估各分支的支持度（Guindon et al., 2010; Hoang et al., 2018）。建树结果均置于 Fig tree 中进行可视化。

2 结果与分析

2.1 三脉水丝梨叶绿体基因组结构和基本特征

三脉水丝梨叶绿体基因组全长 159 375 bp，GC 总含量为 38%（图 1）。其中，大单拷贝区（large single copy，LSC）长 88 067 bp，GC 含量 36.2%；小单拷贝区（small single copy，SSC）18 808 bp，GC 含量 32.5%；单个反向重复区（inverse repeat，IR）长 26 250 bp，GC 含量 43.1%，明显高于 LSC 区和 SSC 区（表 1）。此外，三脉水丝梨叶绿体基因组共编码了 133 个基因，其中包括 8 个 rRNA 基因，37 个 tRNA 基因和 87 个蛋白编码基因和 1 个假基因（#*ycf*1）。按基因功能主要分为 4 大类，即光合作用相关基因、自我复制相关基因、其他基因以及未知功能基因（表 2）。其中，*ndh*B、*rpl*2、*rrn*4.5、*trn*A-UGC、*ycf*15 等 17 个基因有 2 份拷贝，其余基因均为单拷贝基因。此外，三脉水丝梨叶绿体基因组中有 18 个含有内含子的基因，其中 15 个基因（*atp*F、*ndh*A、*ndh*B、*pet*B、*pet*D、*rpl*16、*rpl*2、*rps*16、*rpo*C1、*trn*A-UGC、*trn*G-GCC 、*trn*I-GAU、*trn*K-UUU、*trn*L-UAA、*trn*V-UAC）存在一个内含子、3 个基因（*rps*12、*clp*P、*ycf*3）存在两个内含子。

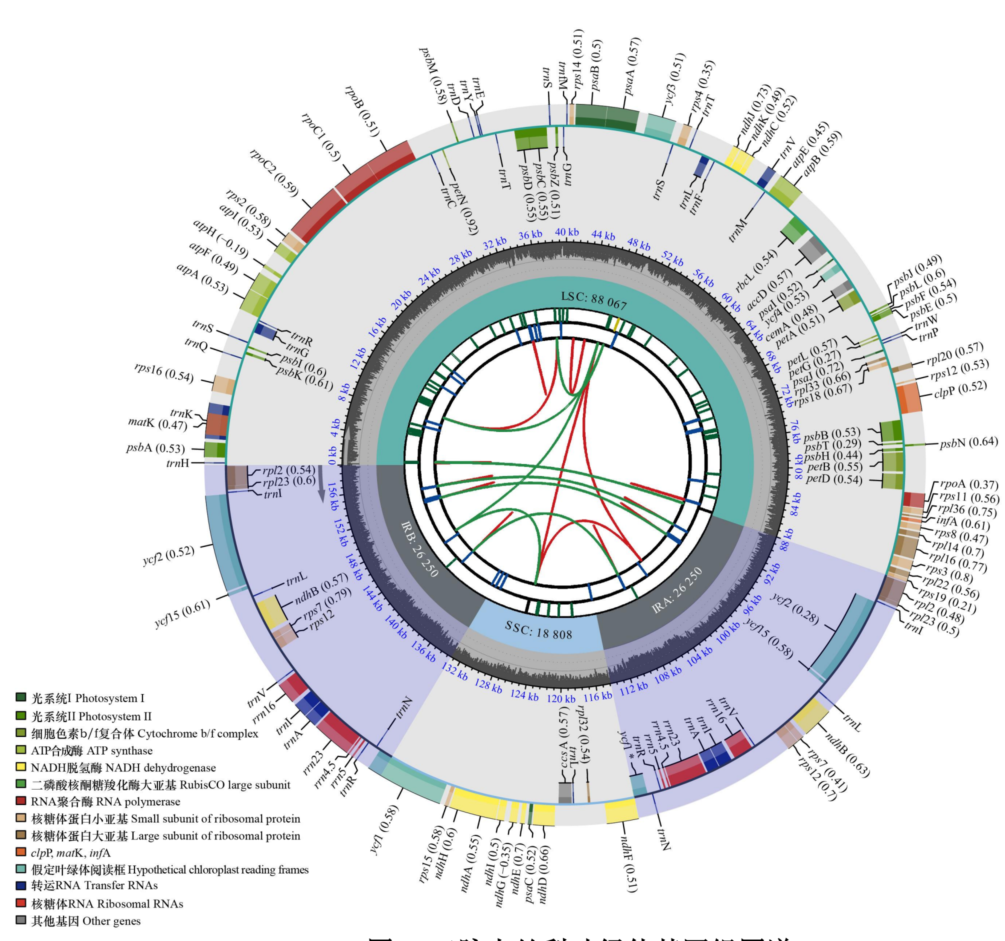


图 1 三脉水丝梨叶绿体基因组图谱

Fig. 1 Genomic map of the chloroplast genome of *Sycopsis triplinervia*

表 1 三脉水丝梨叶绿体基因组不同位置的 GC 含量

Table 1 GC content at different locations in the chloroplast genome of *Sycopsis triplinervia*

名称 Name	大小 Sizes (bp)	开始位置 Start position	结束位置 End position	GC 含量 GC content (%)
大单拷贝区（LSC）	88 067	1	88067	36.2
小单拷贝区（SSC）	18 808	114318	133125	32.5
反向重复区 A（IRA）	26 250	159375	133126	43.1
反向重复区 B（IRB）	26 250	88068	114317	43.1

2.2 重复序列和 SSR 位点分析

利用 REPuter 在三脉水丝梨叶绿体基因组中检测到了 33 个 30~69 bp 的散在重复序列，其中有 1 个反向重复序列，17 个正向重复序列和 15 个回文序列，未检测到互补重复序列。在这些重复序列中有 22 个（包括 14 个正向重复、7 个回文重复和 1 个反向重复序列）位于 *ycf*2、*psa*B、*psa*A、*ycf*3、*ndh*A、*rps*12、*rps*3、*ccs*A、*ndh*D 基因上；3 个回文重复横跨基因（*ycf*3 和 *ndh*A）和基因间隔区（intergenic spacer, IGS）；4 个（1 个正向重复和 3 个回文重复）位于 IGS，4 个（2 个正向重复和 2 个回文重复）位于内含子区域。通过 Tandem repeats finder 软件预测到 39 个串联重复序列，其中长度最小的为 24 bp，最大的为 132 bp，并且串联重复序列主要分布在 IGS 区域（64.10%）。

通过简单重复序列分析，在三脉水丝梨叶绿体基因组中共检测到 82 个 SSRs 位点（图 2），包括 61 个单核苷酸、9 个二核苷酸、3 个三核苷酸、7 个四核苷酸、2 个 5 核苷酸。三脉水丝梨叶绿体基因组中大部分 SSRs 位点（70.73%）位于基因间隔区。此外，A/T 碱基在 SSRs 中出现的频率最高，在 61 个单核苷酸 SSRs 中，有 26 个 SSRs 包含 A 碱基，有 33 个 SSRs 包含 T 碱基。

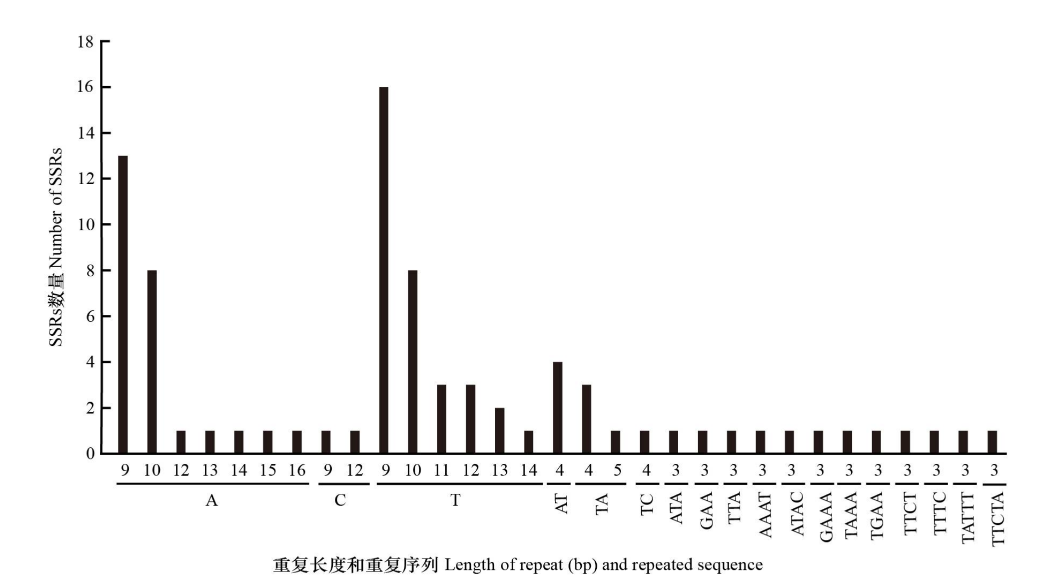


图 2 三脉水丝梨叶绿体基因组中的 SSRs 类型及数量

Fig. 2 Type and number of SSRs in the chloroplast genome of *Sycopsis triplinervia*

表 2 三脉水丝梨叶绿体基因组基因组成

Table 2 Gene annotation of the chloroplast genome of _Sycopsis triplinervia_

基因功能分类 Category of gene	基因分组 Group of gene	基因名 Name of gene
光合作用基因 Genes for photosynthesis	ATP 合成酶 ATP synthase	_atp_A, _atp_B, _atp_E, _atp_F*, _atp_H, _atp_ I
	光系统 II Photosystem II	_psb_A, _psb_B, _psb_C, _psb_D, _psb_E, _psb_F, _psb_H, _psb_I, _psb_J, _psb_K, _psb_L, _psb_M, _psb_N, _psb_T, _psb_Z
	NADH 脱氢酶 NADH dehydrogenase	_ndh_A*, _ndh_B*×2, _ndh_C, _ndh_D, _ndh_E, _ndh_F, _ndh_G, _ndh_H, _ndh_I, _ndh_J, _ndh_K
	细胞色素 b/f 复合体 Cytochrome b/f complex	_pet_A, _pet_B*, _pet_D*, _pet_G, _pet_L, _pet_N
	光系统 I Photosystem I	_psa_A, _psa_B, _psa_C, _psa_I, _psa_J
	二磷酸核酮糖羧化酶大亚基 RubisCO large subunit	_rbc_L
自我复制基因 Self-replication gene	核糖体 RNA Ribosomal RNAs	_rrn_4.5×2, _rrn_5×2, _rrn_16×2, _rrn_23×2
	核糖体蛋白大亚基 Large subunit of ribosomal protein	_rpl_2*×2, _rpl_14, _rpl_16*, _rpl_20, _rpl_22, _rpl_23×2, _rpl_32, _rpl_33, _rpl_36
	RNA 聚合酶 RNA polymerase	_rpo_A, _rpo_B, _rpo_C1*, _rpo_C2
	核糖体蛋白小亚基 Small subunit of ribosomal protein	_rps_2, _rps_3, _rps_4, _rps_7×2, _rps_8, _rps_11, _rps_12** ×2, _rps_14, _rps_15, _rps_16*, _rps_18, _rps_19
	转运 RNA Transfer RNAs	_trn_A-UGC* ×2, _trn_I-CAU ×2, _trn_I-GAU* ×2, _trn_L-CAA×2, _trn_N-GUU×2, _trn_R-ACG×2, _trn_V-GAC×2, _trn_C-GCA, _trn_D-GUC, _trn_E-UUC, _trn_F-GAA, _trnf_M-CAU, _trn_G-GCC*, _trn_G-UCC, _trn_H-GUG, _trn_K-UUU*, _trn_L-UAA*, _trn_L-UAG, _trn_M-CAU, _trn_P-UGG, _trn_Q-UUG, _trn_R-UCU, _trn_S-GCU, _trn_S-GGA, _trn_S-UGA, _trn_T-GGU, _trn_T-UGU, _trn_V-UAC*, _trn_W-CCA, _trn_Y-GUA
其他基因 Other genes	乙酰辅酶 A 羧化酶亚基 Subunit of acetyl-CoA-carboxylase	_acc_D
	C 型细胞色素合成基因 C-type cytochrome synthesis gene	_ccs_A
	包膜蛋白 Envelop membrane protein	_cem_A
	蛋白酶 Protease	_clp_P**
	翻译起始因子 Translational initiation factor	_inf_A
	成熟酶基因 Maturase gene	_mat_K

未知功能基因 Unknown functional gene	假定叶绿体阅读框 Hypothetical chloroplast reading frames	#*ycf*1, *ycf*1, *ycf*2×2, *ycf*3**, *ycf*4, *ycf*15×2

注：*表示基因有一个内含子；**表示基因有两个内含子；#表示假基因；×2 表示具有两个拷贝。

Note: * indicates a gene with one intron; ** indicates a gene with two introns; # indicates a pseudogene; ×2 indicates a gene with two copies.

2.3 密码子偏好性分析

对三脉水丝梨叶绿体基因组的分析中，53 个蛋白质编码基因共有 21 394 个密码子（表 3），其中 2 221 个（10.38%）为编码亮氨酸（leucine, Leu）的密码子，仅有 243 个（1.14%）密码子编码半胱氨酸（cysteine, Cys），表明在三脉水丝梨叶绿体基因组中使用最多的是亮氨酸，使用最少的是半胱氨酸。在所有密码子中，异亮氨酸（isoleucine, Ile）中的 AUU 是使用次数最多的，次数为 885，使用次数最少的是半胱氨酸中的 UGC，次数为 54。3 个终止密码子中 UAA 的使用频率最高（UAA: 50.94%; UAG: 24.53%; UGA: 24.53%）。三脉水丝梨叶绿体基因密码子中，RSCU 值最高的为丙氨酸中的 GCU（1.88）；最低的为丝氨酸中的 AGC（0.33）。有 30 个密码子 RSCU 值大于 1，除 Leu（UUG）对应的密码子外，其余 29 个均以 A/U 结尾，说明三脉水丝梨叶绿体基因组密码子更偏向以 A/U 结尾。

基于 ENC 值建立了高、低表达基因库后，计算出ΔRSCU 值，得到了 28 个三脉水丝梨叶绿体基因组的高表达密码子，其中有 3/7 的密码子以 A/U 结尾。结合筛选得到的高表达密码子和高频密码子，最后得到了 9 个最优密码子（表 3 中带*的密码子）。其中，所有的最优密码子均以 A/U 结尾。

对三脉水丝梨叶绿体基因组密码子第三位碱基进行 PR2-plot 分析（图 3：A），结果表明坐标点主要分布在 A3/(A3+T3)<0.5 和 G3/(G3+C3)>0.5 的区域，总体表现出密码子第三位碱基 G 和 T 的使用频率大于 C 和 A，说明三脉水丝梨叶绿体基因组密码子偏好受到碱基突变和自然选择的影响。从图 3：B 可以看出，基因基本上落在期望曲线下方，说明突变对三脉水丝梨叶绿体基因组密码子偏好性的影响较小。中性绘图分析显示（图 3：C），GC_{12} 与 GC_3 成负相关，相关系数为-0.034，相关性不显著。回归系数为-0.0494，相较而言更接近于 0，表明自然选择对三脉水丝梨叶绿体基因组密码子偏好性的影响最大。

表 3 三脉水丝梨叶绿体基因密码子偏好性分析

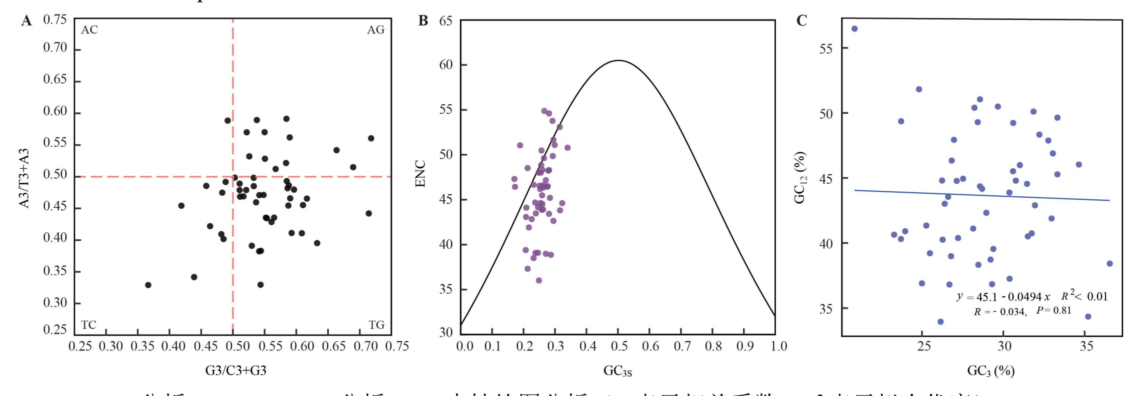
Table 3 Analysis of codon preference of chloroplast genes in *Sycopsis triplinervia*

氨基酸 Amino acid	密码子 Codon	数量 Number	RSCU	氨基酸 Amino acid	密码子 Codon	数量 Number	RSCU
苯丙氨酸 Phe	UUU*	739	1.27	丙氨酸 Ala	GCU	567	1.88
	UUC	422	0.73		GCC	174	0.58
亮氨酸 Leu	UUA	690	1.86		GCA	340	1.13
	UUG	469	1.27		GCG	125	0.41
	CUU*	471	1.27	终止密码子 TER	UAA	27	1.53
	CUC	142	0.38		UAG	13	0.74
	CUA	297	0.80		UGA	13	0.74
	CUG	152	0.41	组氨酸 His	CAU	409	1.54
异亮氨酸 Ile	AUU	885	1.45		CAC	123	0.46
	AUC	391	0.64	谷氨酰胺 Gln	CAA	588	1.53
	AUA	559	0.91		CAG	183	0.47
甲硫氨酸 Met	AUG	504	1.00	天冬酰胺 Asn	AAU*	801	1.57

缬氨酸 Val	GUU	417	1.41		AAC	222	0.43
	GUC	134	0.45	赖氨酸 Lys	AAA	799	1.51
	GUA*	453	1.54		AAG	259	0.49
	GUG	175	0.59	天冬氨酸 Asp	GAU*	714	1.60
丝氨酸 Ser	UCU	448	1.67		GAC	181	0.40
	UCC	253	0.95	谷氨酸 Glu	GAA	871	1.52
	UCA*	324	1.21		GAG	278	0.48
	UCG	154	0.58	半胱氨酸 Cys	UGU	189	1.56
	AGU	338	1.26		UGC	54	0.44
	AGC	88	0.33	色氨酸 Trp	UGG	381	1.00
脯氨酸 Pro	CCU	371	1.65	精氨酸 Arg	CGU	285	1.34
	CCC	164	0.73		CGC	80	0.38
	CCA	262	1.17		CGA	292	1.37
	CCG	102	0.45		CGG	94	0.44
苏氨酸 Thr	ACU	435	1.59		AGA	392	1.84
	ACC	211	0.77		AGG	137	0.64
	ACA*	333	1.22	甘氨酸 Gly	GGU	508	1.35
	ACG	112	0.41		GGC	148	0.39
酪氨酸 Tyr	UAU*	648	1.62		GGA*	610	1.62
	UAC	153	0.38		GGG	241	0.64

注：*表示最优密码子。

Note: * indicates optimal codon.



A. PR2-plot 分析；**B**. ENC-plot 分析；**C**. 中性绘图分析（R 表示相关系数，R^2 表示拟合优度）。

A. Analysis of PR2-plot; **B**. Analysis of ENC-plot; **C**. Analysis of neutrality plot (R indicates correlation coefficient, and R^2 indicates a measure of fitness).

图 3 三脉水丝梨密码子偏好性分析

Fig.3 Analysis of codon preference of chloroplast genes in *Sycopsis triplinervia*

2.4 叶绿体基因组的比较分析

通过在线网站 CPJSdraw 比较分析了三脉水丝梨及其近缘物种的 LSC、SSC 和 IR 区的边界位置，结果如图 4 所示，7 个物种的 IR 区长度较保守（26 218~26 258 bp）。除三脉水丝梨、樟叶假蚊母（*Distyliopsis laurifolia*）和波斯铁木（*Parrotia persica*）外，其余 4 个物种的 LSC/IRB 边界均在 *rps*19 基因内部。樟叶假蚊母的 *ndh*F 基因横跨 IRB/SSC 边界，而其余 6 个物种的 *ndh*F 基因距边界 10~16 bp。7 个物种的 SSC/IRA 边界都位于 *ycf*1 基因上，且

*ycf*1 基因在 IRA 的长度均为 1 017 bp。三脉水丝梨和樟叶假蚊母的 *ycf*1 基因朝 SSC 区分别进入了 4 581 bp 和 4 554 bp，其余 5 个物种均朝 SSC 区进入了 4 560 bp。三脉水丝梨的 *trn*H 基因距离 IRA/LSC 边界最远，为 122 bp，其余 6 个物种 *trn*H 基因与 IRA/LSC 边界的距离差异相对较小（7~25 bp）。

使用 Mauve 插件对三脉水丝梨及 6 个近缘物种的叶绿体基因组进行共线性分析，结果显示 7 个物种基因组具有较高的共线性，基本不存在大片段基因重排和倒置现象，结构比较保守（图 5）。以蚊母树为参照序列，通过 mVISTA 软件对三脉水丝梨及其 6 个近缘物种进行基因组比对，结果显示编码区和 IR 区都更为保守（图 6）。总的来说这 7 个物种的叶绿体基因组不存在大片段缺失的现象，具有较高的相似性。

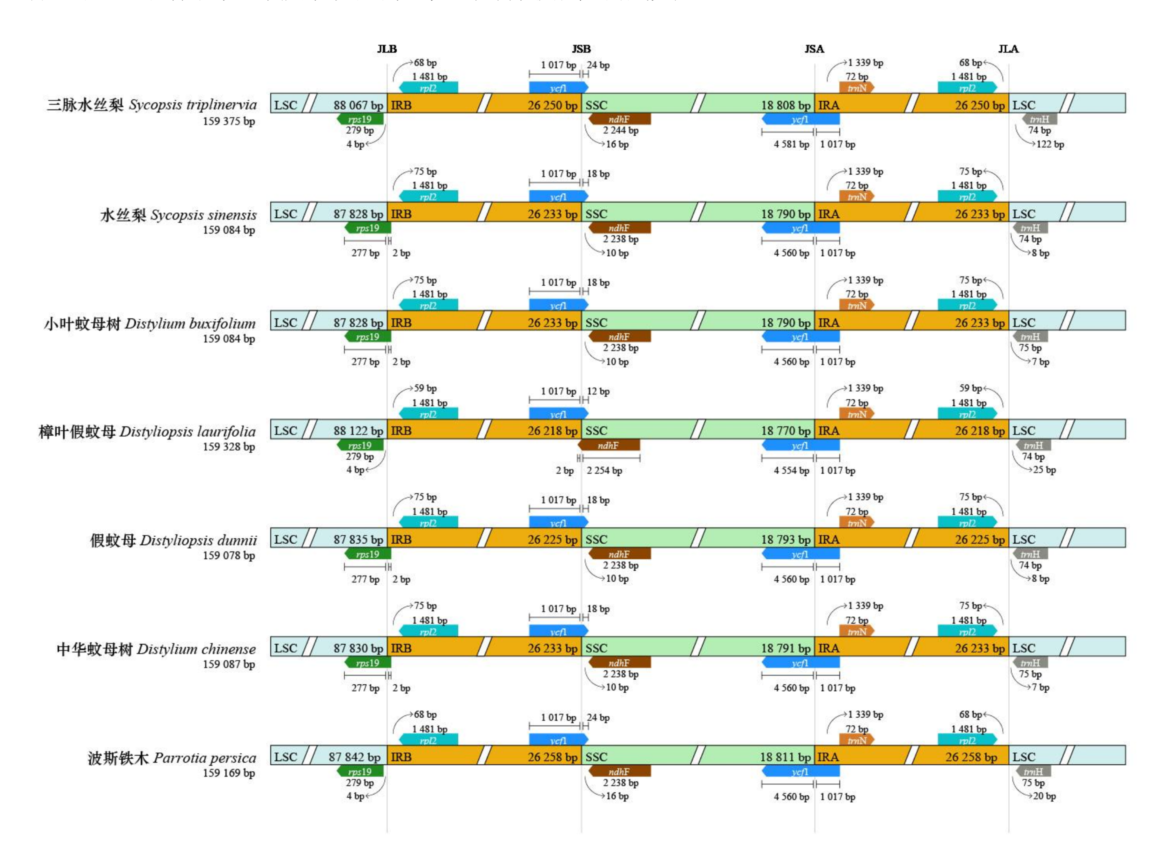


图 4 三脉水丝梨及其 6 个近缘物种叶绿体基因组边界分析

Fig.4 Analysis of chloroplast genome boundaries in *Sycopsis triplinervia* and its six related species

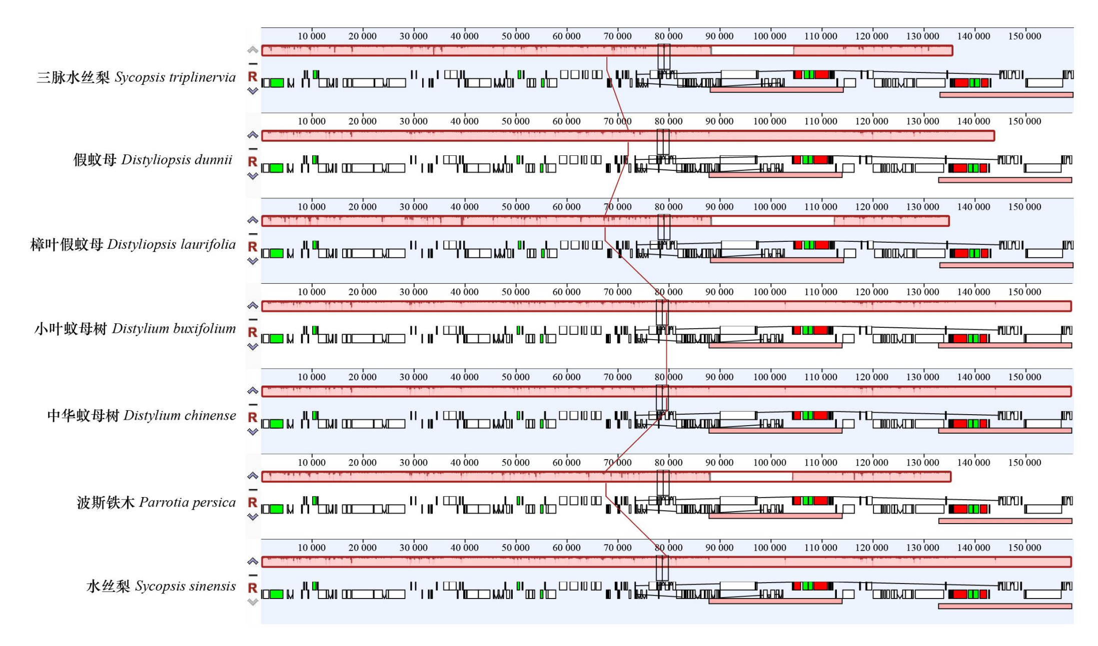


图 5 三脉水丝梨及其 6 个近缘物种叶绿体基因组共线性分析

Fig.5 Collinearity analysis of chloroplast genomes of *Sycopsis triplinervia* and its six related species

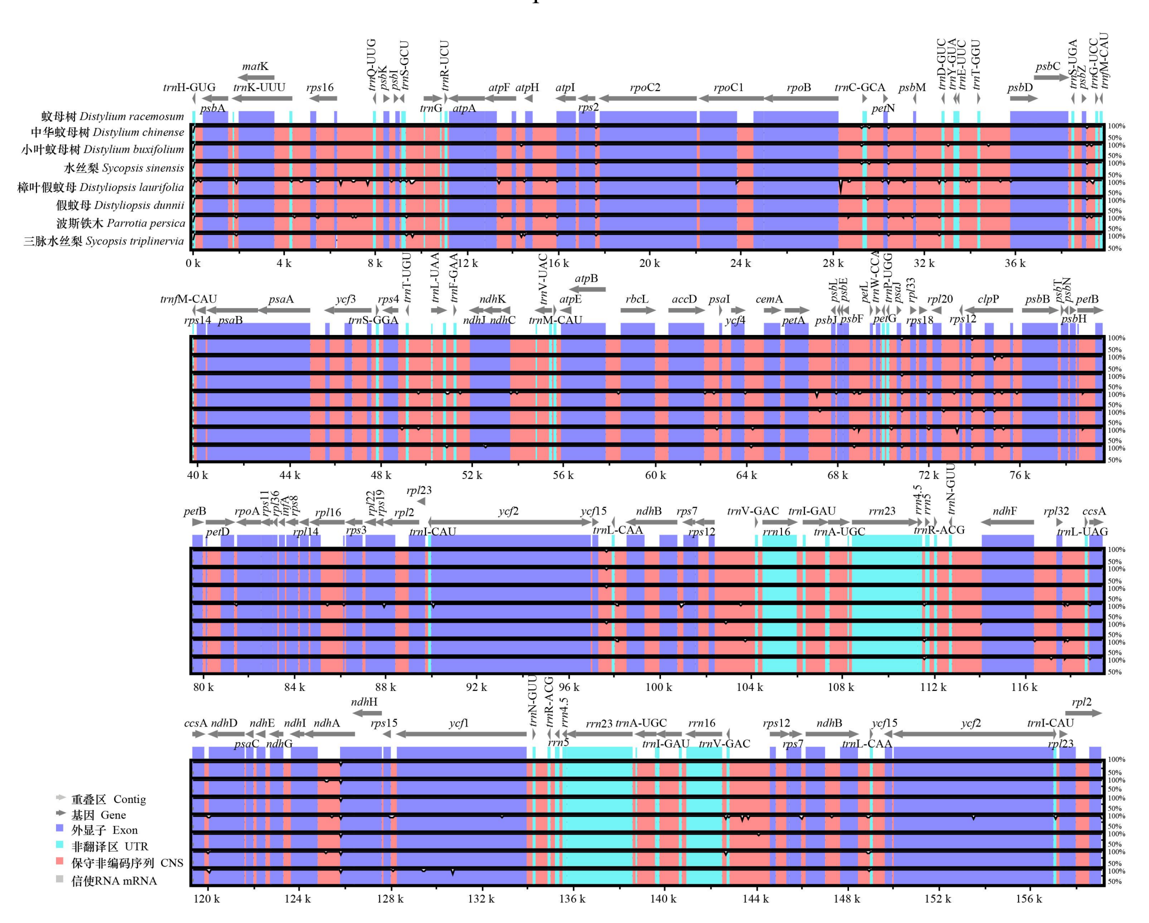
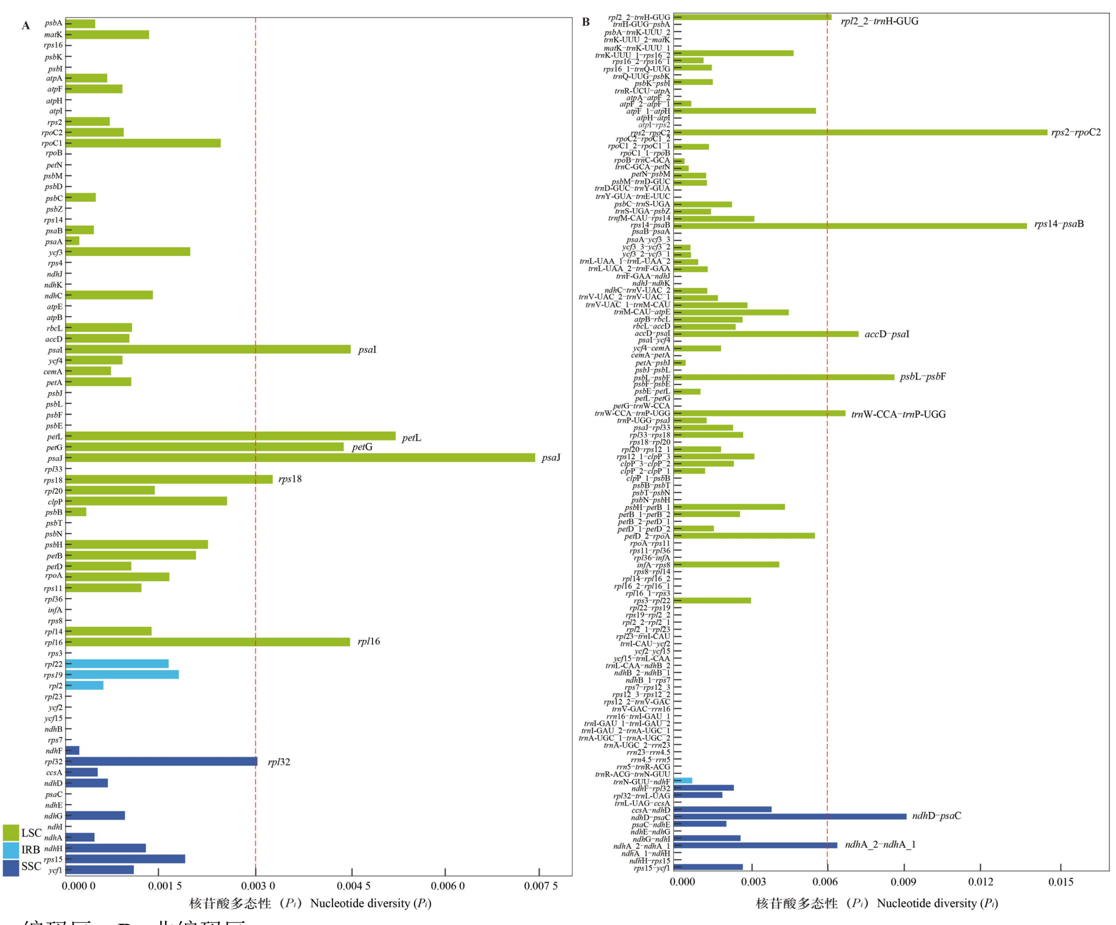


图 6 三脉水丝梨及其 6 个近缘物种叶绿体基因组序列比对

Fig.6 Sequence alignment of chloroplast genomes of *Sycopsis triplinervia* and its six related species

2.5 核苷酸多态性性分析

为确定水丝梨属植物的高突变区域，本研究使用 DnaSP v6.0 软件对水丝梨属三脉水丝梨和水丝梨的 4 条叶绿体基因组的蛋白编码区域和非编码区域分别进行核苷酸多态性（P_i）分析。对 79 个蛋白编码区（重复基因仅计算一次）的分析显示（图 7：A），P_i 值最高的编码区为 *psa*J（0.007 41），其次为 *pet*L（0.005 21）。在蛋白编码区中共发现了 35 处 P_i=0 的高度保守区域。以 P_i=0.003 为阈值，共筛选到 7 个高突变基因，分别为 *psa*J（0.007 41）、*pet*L（0.005 21）、*psa*I（0.004 5）、*rpl*16（0.004 49）、*pet*G（0.004 39）、*rps*18（0.003 27）和 *rpl*32（0.003 03），其中有 6 个高突变基因位于 LSC 区。对非编码区的 P_i 分析显示（图 7：B），P_i 值最高的非编码区为 *rps*2−*rpo*C2（0.014 58），其次为 *rps*14−*psa*B（0.013 78）。在非编码区中共发现了 64 处 P_i=0 的高度保守区域。以 P_i=0.006 为阈值，共筛选到 8 个高变片段，分别为 *rps*2−*rpo*C2（0.014 58）、*rps*14−*psa*B（0.013 78）、*ndh*D−*psa*C（0.009 09）、*psb*L−*psb*F（0.008 62）、*acc*D−*psa*I（0.007 22）、*trn*W-CCA−*trn*P-UGG（0.006 71）、*ndh*A 内含子区域（0.006 39）和 *rpl*2_2−*trn*H-GUG（0.006 17），其中有 6 个位于 LSC 区。水丝梨属蛋白编码区和非编码区的平均 P_i 值分别为 0.000 93 和 0.001 47，非编码区的平均 P_i 值高于蛋白编码区。此外，IR 区 P_i 值显著低于 LSC 区和 SSC 区，水丝梨属植物的 IR 区比 LSC 区和 SSC 区更为保守。



A. 编码区；**B**. 非编码区。

A. Coding regions; **B**. Non-coding regions.

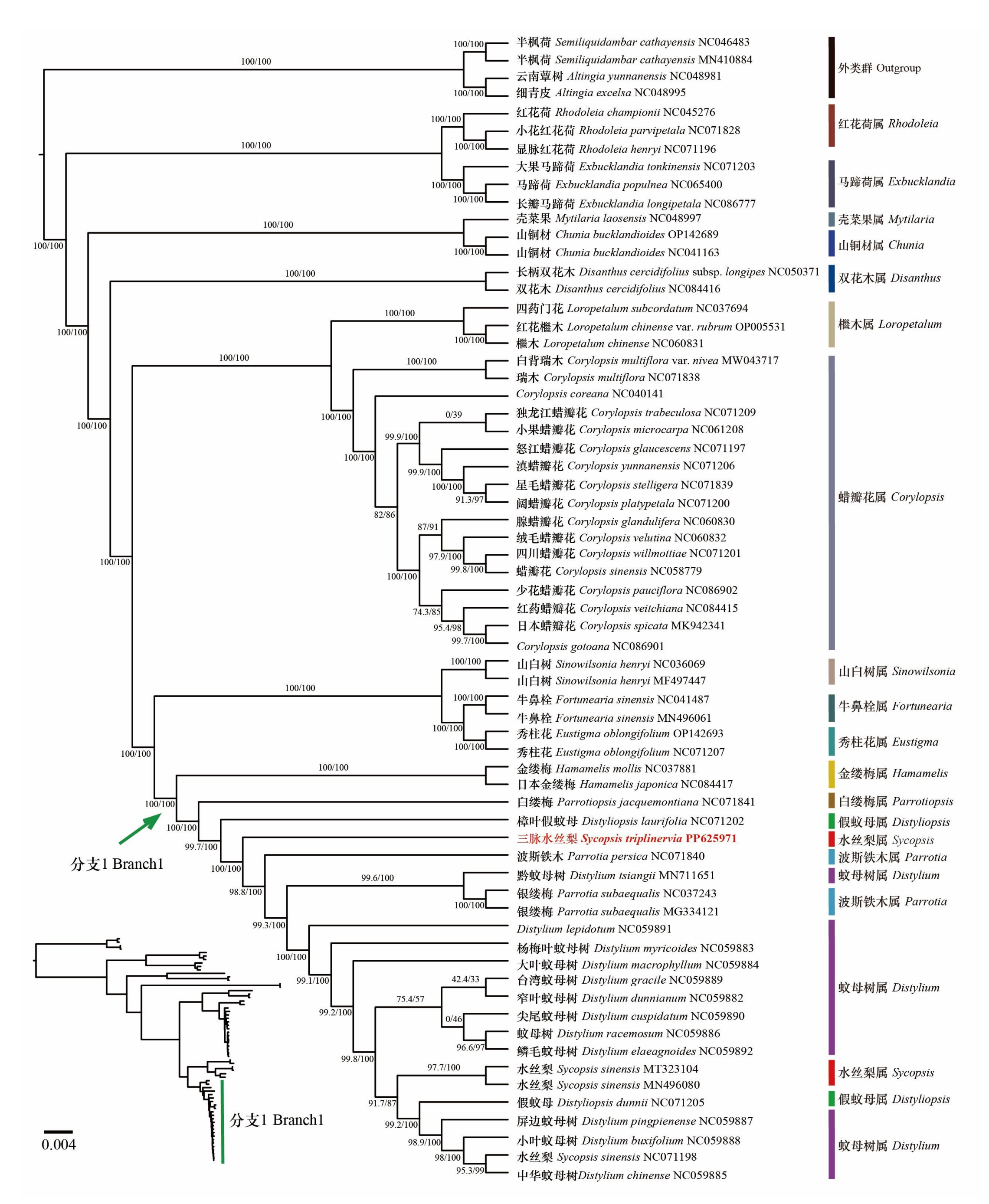
图 7 三脉水丝梨和水丝梨叶绿体基因组核苷酸多样性（P_i）值

Fig.7 Chloroplast genome nucleotide diversity (P_i) values for *Sycopsis triplinervia* and *S. sinensis*

2.6 系统发育分析

选取包括三脉水丝梨在内的 54 种金缕梅科植物的 61 个叶绿体基因组，以蕈树科 3 种植物的 4 个叶绿体基因组为外类群基于最大似然法（ML）构建了金缕梅科叶绿体全基因组系统发育树和共有 CDS 系统发育树（图 8、图 9），两棵系统发育树均显示金缕梅科为得到强烈支持的单系类群（Sh-aLRT/UFboot = 100/100）。金缕梅科的 16 个属中，红花荷属（*Rhodoleia*）和马蹄荷属（*Exbucklandia*）构成的单系分支（Sh-aLRT/UFboot = 100/100）与其余属构成的单系分支（Sh-aLRT/UFboot = 100/100）互为姐妹群。壳菜果属（*Mytilaria*）和山铜菜属（*Chunia*）、双花木属（*Disanthus*）、檵木属（*Loropetalum*）和蜡瓣花属（*Corylopsis*）、山白树属（*Sinowilsonia*）、秀柱花属（*Eustigma*）和牛鼻栓属（*Fortunearia*）分别形成单系分支（Sh-aLRT/UFboot = 100/100）。金缕梅属（*Hamamelis*）、白缕梅属（*Parrotiopsis*）、水丝梨属、假蚊母属、波斯铁木属和蚊母树属 6 个属共同形成了一个得到强烈支持的单系（分支 1，Sh-aLRT/UFboot = 100/100），其中三脉水丝梨与这个分支中的其他类群互为姐妹群，而水丝梨属、波斯铁木属、假蚊母属和蚊母树属均未形成单系分支。

基于叶绿体全基因和共有 CDS 的最大似然树拓扑结构基本一致，不同在于叶绿体全基因组系统树中红药蜡瓣花（*Corylopsis veitchiana*）与日本蜡瓣花（*C. spicata*）和 *C. gotoana* 聚在一起，尖尾蚊母树（*Distylium cuspidatum*）与蚊母树、鳞毛蚊母树（*D. elaeagnoides*）聚在一起；而共有 CDS 系统发育树中则是少花蜡瓣花（*Corylopsis pauciflora*）与日本蜡瓣花和 *C. gotoana* 聚在一支，尖尾蚊母树与窄叶蚊母树（*Distylium dunnianum*）、台湾蚊母树（*D. gracile*）、蚊母树和鳞毛蚊母树一起聚在一支。此外，两棵系统发育树均显示水丝梨的 3 个个体没有形成单系分支，其中一个个体（NC071189）与中华蚊母树是姐妹关系。

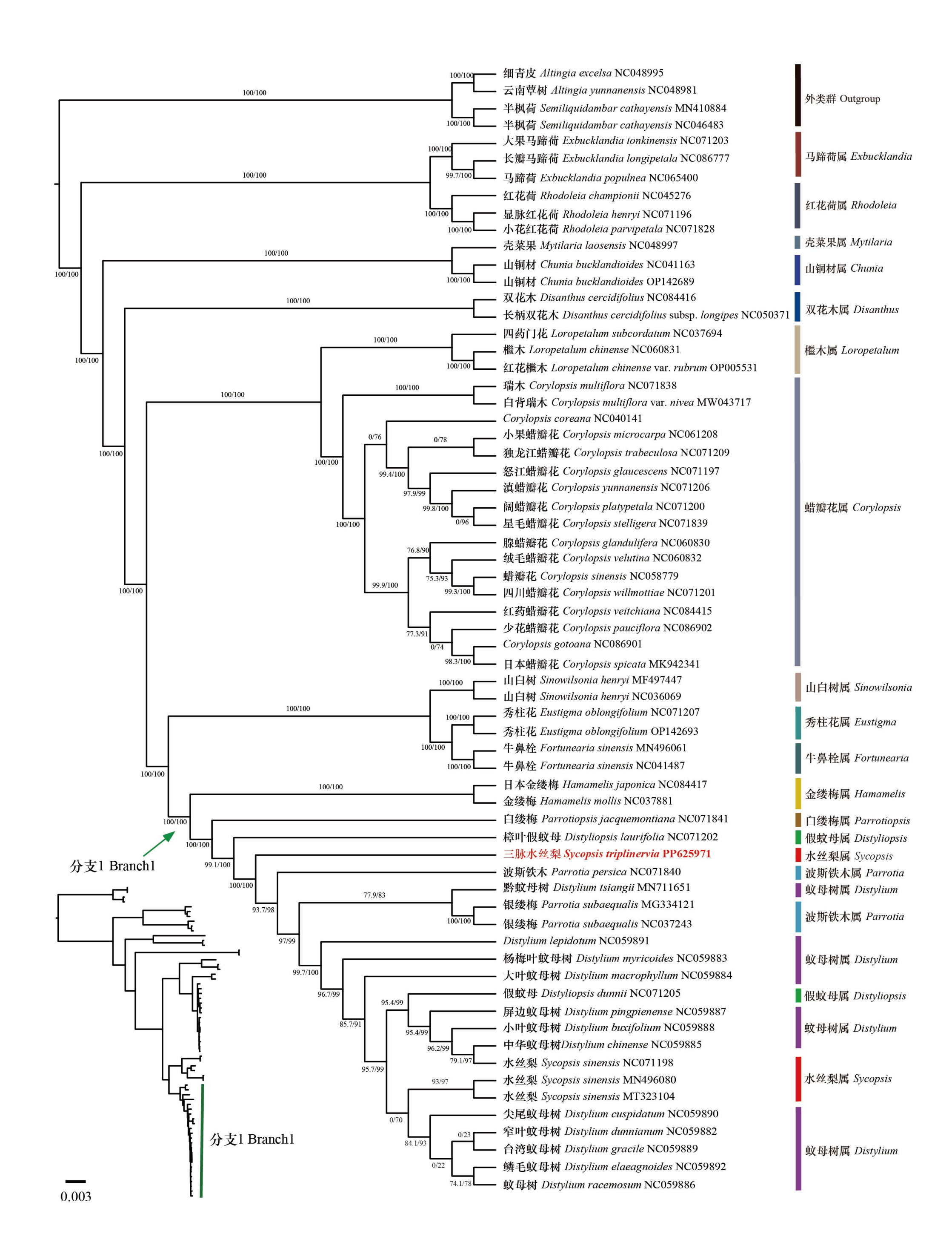


分支支持率分别表示为 SH-aLRT/UFboot。红色加粗字体示三脉水丝梨。

Branch support values are reported as SH-aLRT/UFboot. The red bold font denotes *Sycopsis triplinervia*.

图 8 基于叶绿体全基因组的最大似然树

Fig.8 Maximum likelihood tree based on the complete chloroplast genome



分支支持率分别表示为 SH-aLRT/UFboot。红色加粗字体示三脉水丝梨。

Branch support values are reported as SH-aLRT/UFboot. The red bold font denotes *Sycopsis triplinervia*.

图 9 基于 78 个叶绿体编码基因的最大似然树

Fig.9 Maximum likelihood tree based on 78 chloroplast coding genes

3 讨论与结论

3.1 保守的叶绿体基因组结构

三脉水丝梨叶绿体基因组在基因组的结构、大小、基因的数量、总 GC 含量及 4 个区域的长度和 GC 含量上均与金缕梅科檵木属、蜡瓣花属以及蚊母树属物种基本一致（Dong et al., 2021; Wang et al., 2022），具有高度保守的特点。三脉水丝梨叶绿体基因组的总 GC 含量低于总 AT 含量，且 IR 区的 GC 含量要比 LSC 区和 SSC 区要高。AT 含量高可能是由于同义密码子多以 A/U 结尾，这与进化过程中的自然选择和突变有关（Shimda & Sugiuro, 1991; Clegg et al., 1994; Liu et al., 2019）。此外，rRNA 基因均位于 IR 区且自身具有较高的 GC 含量，因此 IR 区 GC 含量相对更高（Zhang et al., 2012; Wang et al., 2022）。

三脉水丝梨叶绿体基因组共编码 133 个基因，与檵木（*Loropetalum chinense*）和绒毛蜡瓣花（*Corylopsis velutina*）一样都具有假基因#*ycf*1（Wang et al., 2022）。在叶绿体基因组进化过程中存在诸多突变事件，例如插入-缺失（InDels）、替换、倒置和拷贝数变异，而这些突变事件可能导致基因缺失或者假基因化（Bendich et al., 1987; Kumar et al., 2014; Abdullah et al., 2021）。假基因化是一个功能基因变为无功能基因的过程，假基因与功能基因往往有很大的同源性（Wickett et al., 2011; Van et al., 2014）。三脉水丝梨中假基因#*ycf*1 比基因 *ycf*1 要短 4 557 bp，说明其假基因化可能主要是由碱基缺失造成。

3.2 三脉水丝梨叶绿体基因组进化特点

重复序列在分析叶绿体基因组碱基替换、基因组重排以及系统发育等方面有重要的作用（Nie et al., 2012）。三脉水丝梨叶绿体基因组中检测到的 SSRs 位点主要以单核苷酸 A/T 为主，这与金缕梅科其他物种的研究结果一致（Wang et al., 2022）。进一步验证了叶绿体 SSRs 主要是由 A/T 重复所构成，而较少含有 C/G 串联重复的观点（Kuang et al., 2011）。通常，叶绿体基因组具有较高的 AT 含量（Asaf et al., 2018; Liu et al., 2019），而高 A/T 含量的 SSRs 可能是叶绿体基因组 AT 含量高于 GC 含量的潜在原因。

密码子使用偏好是基因组中一个重要的进化特征，在原核生物和真核生物中都广泛存在（Sharp et al., 1988; Wang et al., 2011）。三脉水丝梨叶绿体基因组倾向于使用 UAA 为终止密码子且密码子偏好以 A/U 结尾，其中 9 个最优密码子均以 A/U 结尾，这可能与叶绿体基因组的高 AT 含量有关。密码子偏好的形成受多种因素的影响，其中最主要的是突变和自然选择（Sharp et al., 2010）。经过中性绘图、ENC-plot 等分析发现，自然选择对三脉水丝梨叶绿体基因组密码子偏好性的影响最大。在对胡桃科、菊科等被子植物的研究中也得出了一致的结论，而在拟南芥、玉米等植物中则发现突变才是主要的影响因素（Zhou et al., 2008; Nie et al., 2014；Zeng et al., 2023）。这表明植物对密码子的偏好受到多种因素的影响，而不同的植物受到的主要影响因素也有所差异。

由于缺乏重组以及 IR 区的存在，叶绿体基因组更好地保留了进化历史的特征。IR 区的收缩或扩张是造成叶绿体基因组大小有差异的主要原因，IR 区的扩张可能导致 IR/SC 边界位于附近基因的编码区内，从而导致假基因的出现（Ravi et al., 2008）。本研究发现，三脉水丝梨中 IRA/SSC 边界位于基因 *ycf*1 内部，导致了 IRB/SSC 边界上假基因#*ycf*1 的出现，这可能是金缕梅科的进化过程中发生基因丢失事件所致（Wang et al., 2022）。从叶绿体基因组中筛选到的高变片段，可以用于物种鉴定和系统发育分析。本研究共筛选到 15 个高变区域（例如 *psa*J、*pet*L 和 *rps*14-*psa*B 等），可作为探究水丝梨属系统关系的候选分子标记。

3.3 水丝梨属系统位置及进化关系

叶绿体基因组具有结构保守、进化速率适中和易于测序等特点，被广泛用于植物系统发

育分析（Wolfe et al., 1987; Grevich & Daniell, 2005; Daniell et al., 2016）。本研究基于金缕梅科的叶绿体基因组数据，用最大似然法（ML）构建了金缕梅科系统发育树，重建了金缕梅科属间的系统发育关系，结果表明，金缕梅属、白缕梅属、水丝梨属、假蚊母属、波斯铁木属和蚊母树属 6 个属共同形成了一个得到强烈支持的单系分支，几个属之间具有较近的亲缘关系。

早期基于形态学特征构建的金缕梅科系统框架中，水丝梨属和蚊母树属的系统发育关系比较近（Walker, 1944）。Endress（1970）把原属于水丝梨属的假蚊母、樟叶假蚊母等 4 个物种移出，建立了假蚊母属，并基于有无花瓣、蒴果是否被宿存萼包裹等特征，认为假蚊母属与蚊母树属关系更为密切。而基于 *mat*K 和 ITS 的分子系统学研究也表明水丝梨属、假蚊母属和蚊母树属亲缘关系较近，支持从水丝梨属分出假蚊母属（Li et al., 1999a, b; Xiang et al., 2019）。在形态特征上，水丝梨属、假蚊母属和蚊母树属在花序、叶脉以及蒴果排列方式等方面都具有较大差异，本研究中水丝梨属和假蚊母属均有部分物种镶嵌在蚊母树属中，水丝梨属、假蚊母属和蚊母树属均不形成单系分支。因此，假蚊母属的成立具有一定的合理性，物种间的亲缘关系还需要更深入的研究。但是，Dong 等（2021）基于叶绿体基因组对蚊母树属的系统发育研究则支持蚊母树属为单系类群，而他们的研究中未涵盖水丝梨属和假蚊母属的物种，取样不全可能影响了对蚊母树属是否单系的判断。

我们的结果显示水丝梨属 2 个物种和水丝梨的 3 个个体都没有形成单系分支。这可能是因为水丝梨不同的居群具有较大的遗传分化，或者其中某些个体是通过杂交培育而来。当然，叶绿体基因组序列过于保守也可能导致这一现象。本研究中三脉水丝梨与樟叶假蚊母和波斯铁木都具有较近的亲缘关系，这可能与它们不具有花瓣、嫩枝有星状绒毛的共同特征有关（Zhang et al., 2003）。总之，本研究的结果在一定程度上阐明了水丝梨属、假蚊母属、波斯铁木属和蚊母树属的亲缘关系，但要澄清水丝梨属及其近缘类群的系统关系并进而进行分类学处理，仍需要更为全面的取样，并通过高通量测序技术（如全基因组重测序、转录组测序、基因组浅层测序）获取更多的核基因 DNA 信息位点进行深入的研究。

致谢：感谢中国西南野生生物种质资源库和国家重要野生植物种质资源库提供的三脉水丝梨分子材料以及对本研究的大力支持。

参考文献：

ABDULLAH, MEHMOOD F, HEIDARI P, et al., 2021. Pseudogenization of the chloroplast threonine (*trn*T-GGU) gene in the sunflower family (Asteraceae) [J]. Scientific Reports, 11(1): 21122.

APG IV, 2016. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV [J]. Botanical Journal of the Linnean Society, 181(1): 1-20.

ASAF S, KHAN AL, KHAN MA, et al., 2018. Complete chloroplast genome sequence and comparative analysis of loblolly pine (*Pinus taeda* L.) with related species [J]. Plos One, 13(3): e0192966.

BEIER S, THIEL T, MUENCH T, et al., 2017. MISA-web: a web server for microsatellite prediction [J]. Bioinformatics, 33(16): 2583-2585.

BENDICH AJ, 1987. Why do chloroplasts and mitochondria contain so many copies of their genome? [J]. Bioessays, 6(6): 279-282.

BENSON G, 1999. Tandem repeats finder: a program to analyze DNA sequences [J]. Nucleic Acids Research, 27(2): 573-580.

BOGLE AL, PHILBRICK CT, 1980. A generic atlas of Hamamelidaceous pollens [J].

Contributions from the Gray Herbarium of Harvard University, 210: 29-103.
CAPELLA-GUTIÉRREZ S, SILLA-MARTÍNEZ JM, GABALDÓN T, 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses [J]. Bioinformatics, 25(15): 1972-1973.
CHEN LQ, ZHANG ZR, YANG JB, et al., 2022. Plastid phylogenomic insights into the phylogeny of Convolvulaceae [J] Guihaia, 42(10):1740-1749. [陈丽琼，张志荣，杨俊波，等，2022. 质体系统发育基因组解析旋花科系统发育关系[J]. 广西植物，42(10)：1740-1749.]
CHEN SF, 2023. Ultrafast one-pass FASTQ data preprocessing, quality control, and deduplication using fastp [J]. iMeta, 2(2): e107.
CLEGG MT, GAUT BS, LEARN GH, et al., 1994. Rates and patterns of chloroplast DNA evolution [J]. Proceedings of the National Academy of Sciences, 91(15): 6795-6801.
CORRIVEAU JL, COLEMAN AW, 1988. Rapid screening method to detect potential biparental inheritance of plastid DNA and results for over 200 angiosperm species [J]. American Journal of Botany, 75(10): 1443-1458.
DANIELL H, LIN C-S, YU M, et al., 2016. Chloroplast genomes: diversity, evolution, and applications in genetic engineering [J]. Genome Biology, 17(1): 134.
DARLING ACE, MAU B, BLATTNER FR, et al., 2004. Mauve: Multiple alignment of conserved genomic sequence with rearrangements [J]. Genome Research, 14(7): 1394-1403.
DONG W, LIU Y, XU C, et al., 2021. Chloroplast phylogenomic insights into the evolution of *Distylium* (Hamamelidaceae) [J]. Bmc Genomics, 22(1): 293.
ENDRESS P K, 1989. A suprageneric taxonomic classification of the Hamamelidaceae [J]. Taxon, 38(3): 371-376.
ENDRESS PK, 1970. Die Infloreszenzen der apetalen Hamamelidaceen, ihre grundsätzliche morphologische und systematische Bedeutung [J]. Botanische Jahrbücher für Systematik, 90: 1-54.
FRAZER KA, PACHTER L, POLIAKOV A, et al., 2004. VISTA: computational tools for comparative genomics [J]. Nucleic Acids Research, 32(2): W273-W279.
GREVICH JJ, DANIELL H, 2005. Chloroplast genetic engineering: Recent advances and future perspectives [J]. Critical Reviews in Plant Sciences, 24(2): 83-107.
GUINDON S, DUFAYARD JF, LEFORT V, et al., 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0 [J]. Systematic Biology, 59(3): 307-321.
HARMS H, 1930. Hamamelidaceae [M]. In: ENGLER A and PRANTL K editors. Die Naturlichen Pflanzenfamilien. Vol. 18a. Leipzig: Verlag von Wilhelm Engelmann: 303-345.
HOANG DT, CHERNOMOR O, VON HAESELER A, et al., 2018. UFBoot2: Improving the ultrafast bootstrap approximation [J]. Molecular Biology and Evolution, 35(2): 518-522.
JIN JJ, YU WB, YANG JB, et al., 2020. GetOrganelle: a fast and versatile toolkit for accurate de novo assembly of organelle genomes [J]. Genome Biology, 21(1): 241.
JOHNSON O, 2024. '*Distylium*' from the website Trees and Shrubs Online (treesandshrubsonline.org/articles/distylium/) [EB/OL]. Accessed 2024-08-04.
KATOH K, MISAWA K, KUMA KI, et al., 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform [J]. Nucleic Acids Research, 30(14): 3059-3066.
KUANG DY, Wu H, WangYL, et al., 2011. Complete chloroplast genome sequence of *Magnolia*

kwangsiensis (Magnoliaceae): implication for DNA barcoding and population genetics [J]. Genome, 54(8): 663-673.
KUMAR RA, OLDENBURG DJ, BENDICH AJ, 2014. Changes in DNA damage, molecular integrity, and copy number for plastid DNA and mitochondrial DNA during maize development [J]. Journal of Experimental Botany, 65(22): 6425-6439.
KURTZ S, CHOUDHURI JV, OHLEBUSCH E, et al., 2001. REPuter: the manifold applications of repeat analysis on a genomic scale [J]. Nucleic Acids Research, 29(22): 4633-4642.
LI H, GUO Q, XU L, et al., 2023. CPJSdraw: analysis and visualization of junction sites of chloroplast genomes [J]. Peerj, 11: e15326.
LI JH, 1997. Systematics of the Hamamelidaceae based on morphological and molecular evidence [D]. Durham: University of New Hampshire: 251-273.
LI JH, BOGLE AL, KLEIN AS, 1999a. Phylogenetic relationships in the Hamamelidaceae: evidence from the nucleotide sequences of the plastid gene *mat*K [J]. Plant Systematics and Evolution, 218(3/4): 205-219.
LI JH, BOGLE AL, KLEIN AS, 1999b. Phylogenetic relationships of the Hamamelidaceae inferred from sequences of internal transcribed spacers (ITS) of nuclear ribosomal DNA [J]. American journal of botany, 86(7): 1027-1037.
LIU JQ, 2016. “The integrative species concept” and “species on the speciation way” [J]. Biodiversity Science, 24(9): 1004-1008. [刘建全，2016. “整合物种概念”和“分化路上的物种”[J]. 生物多样性，24(9)：1004-1008.]
LIU X, CHANG E-M, LIU JF, et al., 2019. Complete chloroplast genome sequence and phylogenetic analysis of *Quercus bawanglingensis* Huang, Li et Xing, a Vulnerable Oak Tree in China [J]. Forests, 10(7): 587.
NGUYEN LT, SCHMIDT HA, VON HAESELER A, et al., 2015. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies [J]. Molecular Biology and Evolution, 32(1): 268-274.
NIE X, DENG P, FENG K, et al., 2014. Comparative analysis of codon usage patterns in chloroplast genomes of the Asteraceae family [J]. Plant Molecular Biology Reporter, 32(4): 828-840.
NIE X, LV S, ZHANG Y, et al., 2012. Complete chloroplast genome sequence of a major invasive species, crofton weed (*Ageratina adenophora*) [J]. PLoS ONE, 7(5): e36869.
PEDEN J, 2005. CodonW version 1.4.2 [CP]. Nottingham, UK: University of Nottingham.
PENG Y, YANG L, WEI J, 2020. The complete chloroplast genome of *Sycopsis sinensis* Oliver [J]. Mitochondrial DNA Part B Resources, 5(3): 3002-3003.
POREBSKI S, BAILEY LG, BAUM BR, 1997. Modification of a CTAB DNA extraction protocol for plants containing high polysaccharide and polyphenol components [J]. Plant Molecular Biology Reporter, 15(1): 8-15.
RAVI V, KHURANA JP, TYAGI AK, et al., 2008. An update on chloroplast genomes [J]. Plant Systematics and Evolution, 271(1/2): 101-122.
ROZAS J, FERRER-MATA A, CARLOS SANCHEZ-DELBARRIO J, et al., 2017. DnaSP 6: DNA sequence polymorphism analysis of large data sets [J]. Molecular Biology and Evolution, 34(12): 3299-3302.
SHARP PM, COWE E, HIGGINS DG, et al., 1988. Codon usage patterns in *Escherichia coli*, *Bacillus subtilis*, *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, *Drosophila*

melanogaster and *Homo sapiens*; a review of the considerable within-species diversity [J]. Nucleic Acids Research, 16(17): 8207-8211.
SHARP PM, EMERY LR, ZENG K, 2010. Forces that influence the evolution of codon bias [J]. Philosophical Transactions of the Royal Society B-biological Sciences, 365(1544): 1203-1212.
SHIMDA H, SUGIURO M, 1991. Fine structural features of the chloroplast genome: comparison of the sequenced chloroplast genomes [J]. Nucleic Acids Research, 19(5): 983-995.
TILLICH M, LEHWARK P, PELLIZZER T, et al., 2017. GeSeq-versatile and accurate annotation of organelle genomes [J]. Nucleic Acids Research, 45(W1): W6-W11.
VAN DER BURGT A, JASHNI MK, BAHKALI AH, et al., 2014. Pseudogenization in pathogenic fungi with different host plants and lifestyles might reflect their evolutionary past [J]. Molecular Plant Pathology, 15(2): 133-144.
WALKER EH, 1944. A revision of *Distylium* and *Sycopsis* (Hamamelidaceae) [J]. Journal of the Arnold Arboretum, 25(3): 319-341.
WANG B, YUAN J, LIU J, et al., 2011. Codon usage bias and determining forces in green plant mitochondrial genomes [J]. Journal of Integrative Plant Biology, 53(4): 324-334.
WANG NJ, CHEN SF, XIE L, et al., 2022. The complete chloroplast genomes of three Hamamelidaceae species: Comparative and phylogenetic analyses [J]. Ecology and Evolution, 12(2): e8637.
WICKETT NJ, FORREST LL, BUDKE JM, et al., 2011. Frequent pseudogenization and loss of the plastid-encoded sulfate-transport gene *cys*A throughout the evolution of liverworts [J]. American journal of botany, 98(8): 1263-1275.
WOLFE KH, LI WH, SHARP PM, 1987. Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs [J]. Proceedings of the National Academy of Sciences of the United States of America, 84(24): 9054-9058.
XIANG XG, XIANG KL, ORTIZ RDC, et al., 2019. Integrating palaeontological and molecular data uncovers multiple ancient and recent dispersals in the pantropical Hamamelidaceae [J]. Journal of Biogeography, 46(11): 2622-2631.
ZENG YJ, SHEN LW, CHEN SQ, et al., 2023. Codon usage profiling of chloroplast genome in Juglandaceae [J]. Forests, 14(2): 378.
ZHANG HD, 1979, Flora Reipublicae Popularis Sinicae: Vol. 35, No. 2 [M]. Beijing: Science Press: 36-116. [张宏达，1979. 中国植物志：第 35 卷第 2 分册[M]. 北京：科学出版社：36-116.]
ZHANG TW, FANG YJ, WANG XM, et al., 2012. The complete chloroplast and mitochondrial genome sequences of *Boea hygrometrica*: insights into the evolution of plant organellar genomes [J]. PLoS ONE, 7(1): e30531.
ZHANG ZY, 1999. Notes on the modern classification systems of the Hamamelidaceae [J]. Acta Botanica Yunnanica, 21(1): 1-10. [张志耘，1999. 对金缕梅科现代分类系统的评述 [J]. 云南植物研究，21(1)：1-10.]
ZHANG ZY, ZHANG HD, ENDRESS PK, 2003. Flora of China [M]//WU ZY, RAVEN PH, editors. Vol. 9. Beijing: Science Press; St. Louis: Missouri Botanical Garden Press: 18-42.
ZHOU M, WEI L, XIA L, 2008. Patterns of synonymous codon usage bias in chloroplast genomes of seed plants [J]. Forest Ecosystems, 10(4): 235-242.